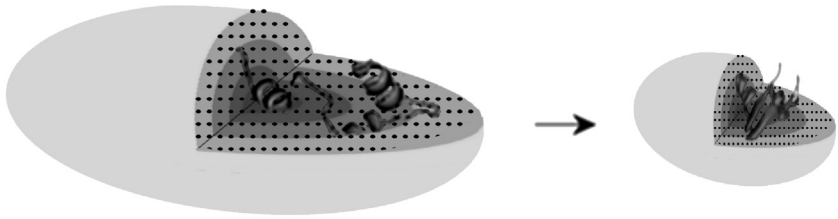
CHAPTER 1

Description of the fuzzy oil drop model

Leszek Konieczny[1], **Irena Roterman**[2]
[1]Chair of Medical Biochemistry, Jagiellonian University – Medical College, Krakow, Poland
[2]Department of Bioinformatics and Telemedicine, Jagiellonian University – Medical College, Krakow, Poland

Contents



This diagram demonstrates the core concept of the fuzzy oil drop (FOD) model, which posits the existence of an external force field generated by the aqueous solvent. The external field guides hydrophobic residues toward the center of the protein body, while hydrophilic residues are instead exposed on its surface. The process continues as the encapsulating ellipsoid capsule shrinks (via changes in σ coefficients of the 3D Gaussian). The gradual increase of gray color visualize the increase of hydrophobicity concentration.

The model has already been presented in detail in numerous publications; nevertheless, we will reintroduce it here in order to provide a theoretical background for further discussions.

The original "oil drop" model, devised many years ago [1] compares the polypeptide to an "oil drop" in the sense that its hydrophobic residues are isolated from the aqueous environment by migrating toward the center of

From Globular Proteins to Amyloids
ISBN: 978-0-08-102981-7
https://doi.org/10.1016/B978-0-08-102981-7.00001-4

the molecule. In parallel, a polar "sheath" emerges to ensure entropically advantageous contact with water. This description may be regarded as a classical discrete qualitative model. As long as hydrophobic residues are entirely isolated from the environment, it appears to work well. If, however, a hydrophilic residue is found to reside at a central location in the protein (or hydrophobic residue exposed on the surface), the model fails to yield accurate results.

In order to mitigate these drawbacks, we have extended the original model, formulating what we refer to as the "fuzzy oil drop" model. This is achieved by introducing a continuous gradient of hydrophobicity, from its maximum value (at the center of the molecule) to near 0 (on the surface). The gradient is mathematically modeled by a 3D Gaussian, which represents the theoretical (or idealized) distribution of hydrophobicity in a perfect protein molecule (i.e. a perfect spherical micelle)—with a prominent central hydrophobic core overlaid by a hydrophilic sheath. Exposure of hydrophilic residues on the surface enables interactions with the aqueous solvent, ensuring solubility. Below we find an intermediate zone, with hydrophobicity increasing along with distance from the surface, depending on the overall volume of the protein body. In a spherical micelle, this gradient is the same in any direction (isotropy), whereas in an elongated globule it may depend on the selected coordinate system axis. The shorter the distance between the surface and the center, the steeper the gradient. The 3D Gaussian is described by three distinct σ coefficients, one for each orthogonal direction. Greater differences between each pair of these values correspond to more elongated globular forms [2,3].

Hydrophobicity is a property of entire amino acid, i.e. a collection of atoms. Thus, for each residue we calculate the position of its so-called effective atom (averaged-out positions of all atoms which comprise that residue). The Gaussian yields a specific value for the location of the effective atom, and this is assumed to represent the theoretical hydrophobicity ascribed to the given residue (T).

The following equation may be used to calculate theoretical hydrophobicity at any point within the ellipsoid capsule:

$$H_i^T = \frac{1}{H_{sum}^T} \exp\left(\frac{-(x_i - \overline{x})^2}{2\sigma_x^2}\right) \exp\left(\frac{-(y_i - \overline{y})^2}{2\sigma_y^2}\right) \exp\left(\frac{-(z_i - \overline{z})^2}{2\sigma_z^2}\right)$$

The point $(\overline{x}, \overline{y}, \overline{z})$is the position of the geometric center of the protein in the 3D coordinate system, when placed in its origin at (0,0,0), these values

become 0. The protein should be rotated, making the line linking longest distance between two effective atoms in the molecule coaxial with (say) X-axis. It is then rotated around the X-axis to make the line linking the two most distant positions of the projections of effective atoms on the (say) YZ plane coaxial with Y-axis. Three parameters σ_x, σ_y, σ_z represent standard deviations of the size of the protein, equal to 1/3 of the highest absolute values of x-coordinate, y-coordinate and z-coordinate respectively (according to the 3-sigma rule). The normalizing coefficient H_{sum}^T represents the sum of all H_i^T values of amino acids of the protein, making the H_i^T value normalized. The only input information for the theoretical distribution is a geometrical term concerning the full protein, i.e. the size of the ellipsoid "drop," containing the protein, and characterized by σ_x, σ_y, σ_z. Traditionally, value of the Gauss function is interpreted as a theoretical idealized hydrophobicity density at given point.

Of course, a real protein is not expected to conform to this model with perfect accuracy. Thus, we calculate the actual (observed) hydrophobicity for each amino acid, which depends on its own intrinsic hydrophobicity (according to any generally accepted scale) as well as on interactions with its neighbors. Following [4], we calculate the hydrophobic interaction assuming a cutoff distance of 9 Å for hydrophobic interactions.

$$H_i^o = \frac{1}{H_{sum}^o} \sum_j \begin{cases} \left(H_i^r + H_j^r\right)\left(1 - \frac{1}{2}\left(7\left(\frac{r_{ij}}{c}\right)^2 - 9\left(\frac{r_{ij}}{c}\right)^4 + 5\left(\frac{r_{ij}}{c}\right)^6 - \left(\frac{r_{ij}}{c}\right)^8\right)\right), & for r_{ij} \le c \\ 0, & for r_{ij} > c \end{cases},$$

Where H_i^o denotes the experimentally observed (index O) hydrophobic density in particular point (position of effective atom of *i-th* residue) which collects the hydrophobic interaction in distance dependent form as given in the formula with the cutoff distance (c) assumed according to original work 9 Å [4]. The denominator H_{sum}^o (sum of all H_i^o) makes the value in normalized form. H_i^r and H_j^r express the intrinsic hydrophobicity of *i-th* an *j-th* residues, which can be taken according to arbitrarily selected scale [5,6]. The scale presented in Ref. [6] was taken for calculation in the work discussed here.

Interactions with neighbors may either increase or lower the effective hydrophobicity of each residue—this is reflected by its corresponding observed hydrophobicity value.

Comparison of T and O (following normalization) enables us to unambiguously determine the accordance/discordance between both distributions, either for the protein as a whole, or for selected fragments. We may even point to specific residues as either accordant or discordant versus the theoretical model (O_i vs. T_i).

Given both distributions, it is possible to perform a quantitative comparison by applying Kullback-Leibler's [7] divergence entropy formula:

$$D_{KL}(P \parallel Q) = \sum_i P(i) \log_2 \frac{P(i)}{Q(i)}$$

Where $P(i)$ denotes the observed probability (hydrophobicity density) localized on i-th residue—in this paper called O_i—observed and $Q(i)$ denotes the expected (target distribution) hydrophobicity localized on the same residue—in this paper called T_i—theoretical one—corresponds to the distance between O and T, the latter of which is regarded as the reference.

D_{KL} expresses the formal "distance" between both distributions (T and O). However, since it is constitutes a measure of entropy, the value of D_{KL} cannot be interpreted on its own—a second reference model must be provided. Since T simulates a "perfect" centric hydrophobic core, we may add a reference distribution which lacks any concentration of hydrophobicity at any point in the protein body. This type of distribution—called the unified distribution (R)—assigns hydrophobicity of 1/N to each residue (N being the number of residues in the chain). It represents the status deprived of any form of hydrophobicity differentiation in protein body.

When considering O and R, the value of D_{KL} tells us to what degree the observed distribution approximates the unified distribution. Comparing both values (for O/T and O/R) provides a description of the protein's status: when $O/T < O/R$, the observed distribution is aligned with the theoretical distribution, and therefore the protein may be assumed to contain a hydrophobic core. In the opposite case—$O/T > O/R$—the protein lacks a prominent core.

By applying the 3D Gaussian model and calculating divergence entropy, we obtain a fine-grained description of the protein's status—a procedure which would not be possible under the original "oil drop" model.

In order to avoid having to deal with two distinct values of D_{KL}, we compute another parameter referred to as relative distance (RD):

$$\mathrm{RD} = \mathrm{O|T}/(\mathrm{O|T} + \mathrm{O|R})$$

Where

O|T denotes according to D_{KL} definition

$$\mathrm{O|T} = \sum_{i=1}^{N} \mathrm{O}_i \log_2(\mathrm{O}_i/\mathrm{T}_i)$$

And

$$\mathrm{O|R} = \sum_{i=1}^{N} \mathrm{O}_i \log_2(\mathrm{O}_i/\mathrm{R}_i)$$

RD (T-O-R) expresses the relation between O and two other distributions—T and R—treated as edge cases. Unlike D_{KL}, this value is independent of the length of the chain and may be used to characterize any protein.

In our search for the causes of discordance between T and O, we have also introduced another type of reference distribution which expresses the intrinsic hydrophobicity of each residue in the input chain. This distribution is labeled H and may be swapped in for R to determine whether the observed structure is dominated by the intrinsic properties of its component residues—as expressed by the corresponding value of RD (T-O-H):

$$\mathrm{RD} = \mathrm{O|T}/(\mathrm{O|T} + \mathrm{O|H})$$

Where O|H according to D_{KL} definition:

$$\mathrm{O|H} = \sum_{i=1}^{N} \mathrm{O}_i \log_2(\mathrm{O}_i/\mathrm{H}_i)$$

High values (above 0.5) of RD suggest that no hydrophobic core is present, and additionally that the observed conformation is driven by the "selfish" properties of each residue rather than by the synergistic tendency to produce a shared core. This is why we distinguish RD for (T-O-R) and RD for (T-O-H) relations.

The presented assumptions, when applied to specific proteins, allow us to validate the model as a whole. A full description of the fuzzy oil drop model also needs to acknowledge the correlations between each pair of distributions (T, O and H), expressed as three distinct correlation coefficients

(TvO, TvH and OvH). These coefficients are based on values obtained for selected fragments or for short overlapping fragments of the chain (5 residues each, produced by a moving frame algorithm) and are particularly useful in identifying discordant fragments. Comparing all three coefficients points to specific locations where discordances occur, and also explains their character (by highlighting the causative factor).

In summary, it should be noted that the fuzzy oil drop model is a reflection of a synergistically generated hydrophobic core, which depends on cooperation between residues belonging to the polypeptide chain. Where such cooperation does not occur, the folding process is driven by the intrinsic properties of each residue, which may be regarded as "selfish" action. Under these conditions, the protein cannot reach a globular conformation and alternative structural patterns emerge—including, in some cases, amyloid forms. The specific goal of this study is to suggest a certain path and a mechanism explaining the changes which accompany amyloidogenesis.

Fig. 1.1 provides a graphical depiction of the presented model, along with its interpretation.

Fig. 1.1A visualizes the theoretical distribution, modeled by the 3D Gaussian (T—blue), observed one (O-red) and uniform (R—green). The status of the O distribution in the RD scale is described as discordant versus T distribution. The RD (T-O-R) value equal to 0.693 suggests the closeness versus the uniform distribution. This is why the O distribution is interpreted as lacking the uni-centric hydrophobic core.

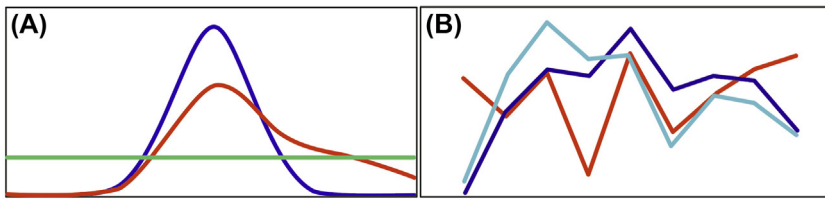


Fig. 1.1 Graphical representation of fuzzy oil drop model parameters reduced to a single dimension for simplicity. (A) theorized Gaussian distribution (T—blue (black in print version)), observed distribution (O—red (gray in print version)) and uniform distribution (R—green (light gray in print version)). (B) theoretical (idealized) hydrophobicity distribution—blue (black in print version), observed—red (gray in print version) and intrinsic hydrophobicity according to the sequence—light blue(light gray in print version). This calculation describes the status of certain polypeptide chain fragment limited to 9 residues.

Fig. 1.1B shows the theoretical distribution (T—blue) for a selected fragment of the chain, while the corresponding observed distribution (O—red) and the intrinsic hydrophobicity values for each residue belonging to the analyzed fragment is shown as H—light blue. The status of discussed chain fragment is expressed by RD (T-O-H) value equal to 0.586. It suggests that the O distribution does not follow the T distribution being dependent on intrinsic hydrophobicity of residues present in the polypeptide chain fragment. The characteristics of selected fragment defines its status in respect to the structural unit (chain or domain).

For the sake clarity, our presentation is limited to a single dimension; however all computations are conducted for the three-dimensional structure (described by the previously mentioned 3D Gaussian).

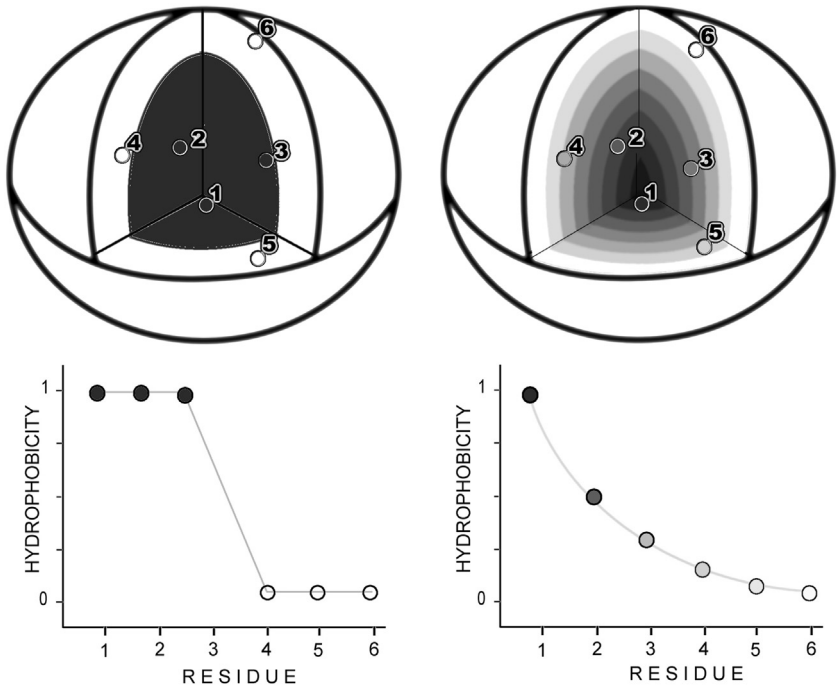


Fig. 1.2 Distribution of hydrophobicity in a protein molecule. Instead of black-white discrete status the continuous one is introduce in fuzzy oil drop model. The positions of residues harmonize their intrinsic hydrophobicity with the expected level of hydrophobicity in protein body. Left -—discrete oil drop model—two layers: black—hydrophobic core; white—hydrophilic shell. Right—fuzzy oil drop model—distribution of hydrophobicity asserted by the model.The bottom profiles express the status of each residue.

The observations described in Refs. [8–11] are treated as supporting the definition of the presented model.

Fuzzy oil drop versus oil drop model

To summarize the description of fuzzy oil drop model the graphic presentation (which appeared in few papers) is recalled. It visualizes the progress from "oil drop"—which is of discrete form to "fuzzy oil drop" of continuous character.

The applicability of fuzzy oil drop model is seen well when the residues do not obey the expected hydrophobicity distribution. Instead of two discrete forms the continuous interpretation is possible allowing measurement of the discordance status (Fig. 1.3).

The interpretation of profiles expressed by fuzzy oil drop model allows distinguishing of local accordance: residues 1, 5, 6, 7 and 9 seem to represent

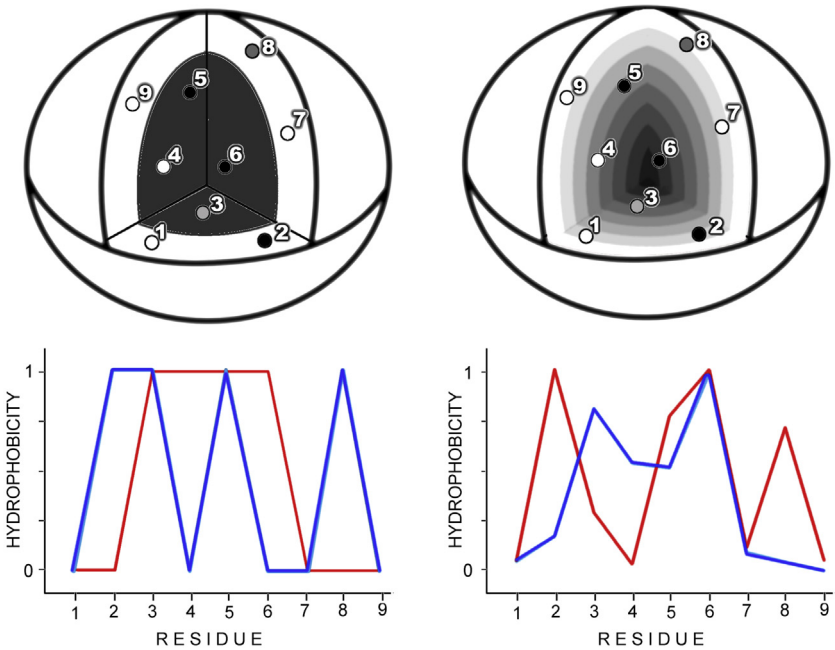


Fig. 1.3 Discordant distribution of hydrophobicity. The positions of some residues cause collision between their status in protein body and the expected level of hydrophobicity. Left—discrete model with distribution of residues not corresponding to the two-layer structure together with profile visualizing the distribution. Right—distribution of hydrophobicity in the continuous model: theoretical (T—blue (gray in print version)) and observed (O—red (dark gray in print version)) together with profile visualizing the distribution. T—blue line (gray in print version), O—red line (gray in print version).

the status accordant with expectation. The residues 3, 4 and 8 can be identified as introducing local discordance.

Amyloids as seen from the perspective of the FOD model

Applying the fuzzy oil drop model to various classes of proteins reveals numerous instances where observed structures closely correspond to theoretical predictions. This is particularly true for individual domains, which in majority contain monocentric cores [8]. Local discordances are frequently associated with biological function—excess hydrophobicity on the protein surface indicates potential complexation sites [9,10], while local hydrophobicity deficiencies often indicate the presence of ligand (or substrate) binding cavities [11].

A particularly interesting case of discordance is observed in amyloids [12]. The observed distribution of hydrophobicity in such structures diverges greatly from the corresponding theoretical distribution, in favor of linear propagation of alternating bands of high and low hydrophobicity parallel to the fibril's axis. Such linear propagation is, by definition, unlimited, enabling unrestricted elongation of the amyloid fibril. In contrast, a prominent hydrophobic core surrounded by a hydrophilic "shell" produces a soluble, globular protein. Fig. 1.4 depicts the structural differences between a globular molecule and a linearly propagating fibril.

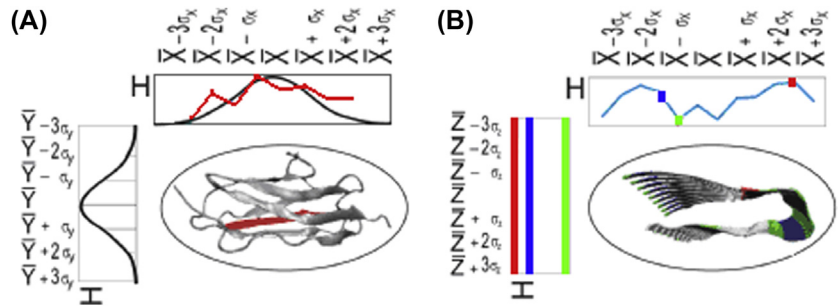


Fig. 1.4 Comparison of a monocentric globular molecule (A) where the observed distribution (red line (gray in print version) representing the status of the fragment distinguished as red in 3D presentation) closely corresponds to the theoretical model (black), and an amyloid (B), which exhibits linear propagation (distribution along the Z-axis perpendicular vs. the plane of the picture of alternating bands of high and low hydrophobicity) [12]. See Ref. [12] for a more detailed introduction to the model. The colors in 3D presentation in B visualize the linear propagation of low/high hydrophobicity (color of points respectively).

It furthermore appears that the linear (not uniform) distribution of hydrophobicity observed in amyloids is closely aligned to the intrinsic distribution (H). One shall distinguish the uniform distribution (called as R) which is isotropic—the constant hydrophobicity is expected in any point of protein body. While the linear distribution expresses the formation of linear bands of different hydrophobicity along and parallel to long axis of the molecule.

In this context, amyloid transformation can be interpreted as optimization of hydrophobic interactions in line with the intrinsic properties of each residue. Minimizing the influence of the aqueous environment may favor such transformation—in contrast to the folding of globular proteins, which depends strongly on active interaction with the aqueous environment (resulting in internalization of hydrophobic residues and exposure of hydrophilic residues on the protein surface) [13-16].

The structural diversity of proteins will be illustrated on the example of titin (good correspondence between T and O) and amyloid Aβ4 (classic linear propagation of alternating bands). Transthyretin appears to represent the middle ground between these two boundary cases. The possible mechanism of transthyretin amyloid transformation is discussed in details in Chapter 12.

References

[1] Kauzmann W. Some factors in the interpretation of protein denaturation. Advances in Protein Chemistry 1959;14:1−63. https://doi.org/10.1016/s0065-3233(08)60608-7.
[2] Konieczny L, Brylinski M, Roterman I. Gauss-function-Based model of hydrophobicity density in proteins. In Silico Biology 2006;6(1−2):15−22.
[3] Roterman-Konieczna I. Protein folding in silico. 2012. https://doi.org/10.1533/9781908818256.
[4] Levitt M. A simplified representation of protein conformations for rapid simulation of protein folding. Journal of Molecular Biology 1976;104(1):59−107. https://doi.org/10.1016/0022-2836(76)90004-8.
[5] Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of a protein. Journal of Molecular Biology 1982;157(1):105−32. https://doi.org/10.1016/0022-2836(82)90515-0.
[6] Kalinowska B, Banach M, Konieczny L, Roterman I. Application of divergence entropy to characterize the structure of the hydrophobic core in DNA interacting proteins. Entropy 2015;17(3):1477−507. https://doi.org/10.3390/e17031477.
[7] Kullback S, Leibler RA. On information and sufficiency. The Annals of Mathematical Statistics 1951;22(1):79−86. https://doi.org/10.1214/aoms/1177729694.
[8] Sałapa K, Kalinowska B, Jadczyk T, Roterman I. Measurement of hydrophobicity distribution in proteins - non-redundant protein data bank. Bio-Algorithms and Med-Systems 2012;8(2):327−38. https://doi.org/10.2478/bams-2012-0023.
[9] Banach M, Konieczny L, Roterman I. Use of the "fuzzy oil drop" model to identify the complexation area in protein homodimers. In: Roterman-Konieczna I, editor. Protein folding. Silico. Oxford, Cambridge, Philadelphia, New Delhi: Woodhead Publishing (currently Elsevier); 2012. p. 95−122.

[10] Banach M, Konieczny L, Roterman I. Can the structure of hydrophobic core determine the complexation site? In: Roterman-Konieczna I, editor. Identification of ligand binding site and protein-protein interaction area. Springer; 2013. p. 41–54.
[11] Banach M, Konieczny L, Roterman I. Recognition of ligand binding binding site. In: Roterman-Konieczna I, editor. Protein folding in Silico. Woodhead publishing (currently Elsevier; 2012. p. 79–93.
[12] Roterman I, Banach M, Konieczny L. Application of the fuzzy oil drop model describes amyloid as a ribbonlike micelle. Entropy 2017;19(4):167. https://doi.org/10.3390/e19040167.
[13] Manoharan VN. Pinned down. Nature Materials 2015;14(9):869–70. https://doi.org/10.1038/nmat4400.
[14] Manoharan VN. Colloidal matter: packing, geometry, and entropy. Science 2015; 349(6251). https://doi.org/10.1126/science.1253751. 1253751–1253751.
[15] Macias-Romero C, Nahalka I, Okur HI, Roke S. Optical imaging of surface chemistry and dynamics in confinement. Science 2017;357(6353):784–8. https://doi.org/10.1126/science.aal4346.
[16] Schutzius TM, Jung S, Maitra T, Graeber G, Köhme M, Poulikakos D. Spontaneous droplet trampolining on rigid superhydrophobic surfaces. Nature 2015;527(7576): 82–5. https://doi.org/10.1038/nature15738.